DOCUMENT RESUME

ED 476 867 TM 034 967

| | |
|---|---|
| AUTHOR | Kang, Taehoon; Cohen, Allan S. |
| TITLE | A Mixture IRT Model Analysis of Ethnic Group DIF. |
| PUB DATE | 2003-04-23 |
| NOTE | 33p.; Paper presented at the Annual Meeting of the National Council of Measurement in Education (Chicago, IL, April 22-24, 2003). |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | EDRS Price MF01 Plus Postage. PC Not Available from EDRS |
| DESCRIPTORS | Ethnicity; Item Bias; Item Response Theory; *Racial Differences; *Student Characteristics; *Test Items |
| IDENTIFIERS | *Item Bias Detection |

ABSTRACT

A number of methods exist for detection of differential item functioning (DIF), but these methods tell us little about the causes of DIF. DIF is typically defined based on a relationship with some manifest group characteristic, such as gender or ethnicity, which is only weakly associated with DIF. What is lacking is a method that will lead to understanding why some examinees respond differently to an item and others do not. In this study, item response patterns on a set of 1996 National Assessment of Educational Progress grade 8 State Mathematics Assessment items for ethnic group DIF were analyzed in two different samples. In one sample, responses from Caucasian and African American examinees were compared, and in a second sample, responses from Caucasian and Hispanic examinees were compared. The patterns of DIF in these two samples were used to define two separate mixture item response theory models in an attempt to determine whether the manifest characteristics of ethnic group membership was useful in identifying those examinees who were actually advantaged or disadvantaged by the item(s) in either sample. Comparisons of background and attitude variables of these examinees revealed some clear differences between manifest and latent groups. (Contains 4 figures, 7 tables, and 21 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 476 867

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

T. Kang

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

# A Mixture IRT Model Analysis of Ethnic Group DIF

Tachoon Kang and Allan S. Cohen

University of Wisconsin–Madison

April 23, 2003

TM034967

Running Head: Mixture Model Analysis of DIF

Paper Presented at the:
Annual Meeting of the National Council of Measurement in Education
Chicago, IL- April 2003

BEST COPY AVAILABLE

**Abstract**

A number of methods exist for detection of DIF, but these methods tell us little about the causes of DIF. DIF is typically defined based on a relationship with some manifest group characteristic, such as gender or ethnicity, which is only weakly associated with DIF. What is lacking is a method that will lead to understanding why some examinees respond differentially to an item and others do not. In this study, we analyzed item response patterns on a set of 1996 NAEP Grade 8 State Mathematics Assessment items for ethnic group DIF in two different samples. In one sample we compared responses from Caucasian and African-American examinees and in a second sample, we compared responses from Caucasian and Hispanic examinees. We then used the patterns of DIF in these two samples to define two separate mixture item response theory models in an attempt to determine whether the manifest characteristic of ethnic group membership was useful in identifying those examinees who were actually advantaged or disadvantaged by the item(s)in either sample. Comparisons of background and attitude variables of these examinees revealed some clear differences between manifest and latent groups.

*Index Terms: DIF, Item Response Theory, Mixture Models, Markov Chain Monte Carlo Estimation*

## A Mixture Model Analysis of Ethnic Group DIF

A number of methods exist for detection of differential item functioning (DIF), but once a DIF item has been identified, little is known about the examinees for whom the item actually functions differentially. DIF is typically investigated by associating some manifest examinee characteristic(s), such as ethnicity or gender, with differential performance on the item. Unfortunately, this approach does not explain why examinees in the different groups responded differentially to the item. The reason is that these DIF detection methods define examinee characteristics of interest but not the dimension(s) causing the DIF. Consequently, when differences are found, it is not so readily apparent who is primarily being advantaged or disadvantaged by the DIF items (Cohen & Bolt, 2002). In fact, the variable(s) used to distinguish groups (e.g., gender, ethnicity, or socio-economic status) may not be the most useful in understanding who is being unfairly disadvantaged by the item. Identifying the characteristics of these examinees should help us understand more about what may have caused the DIF.

A common explanation for DIF is that differentially functioning items actually measure nuisance dimension(s) unrelated to the dimension intended to be measured by the test (Ackerman, 1992; Roussos & Stout, 1996). DIF is typically observed when these nuisance dimensions are in some way associated with manifest examinee characteristics. In this study, we consider the use of an IRT mixture model for investigating nuisance dimensions associated with ethnic group-related differences in item functioning. Here, we deal with three ethnic groups, Caucasian, African-American, and Hispanic. Specifically, we begin by identifying DIF items on the basis of membership in one of these groups, but assume that the actual respondents advantaged or disadvantaged by the items are better regarded as latent classes in the data. We assume, in other words, that within each of these ethnic groups there are some examinees who in reality respond more like members from the other groups. Next, by identifying other distinguishing

BEST COPY AVAILABLE

characteristics of the members of these latent classes (besides ethnicity), we can provide useful information about where and for what kinds of examinees DIF items function differentially.

Methods for detection of DIF will usually result in identification of some differentially functioning items. Within the context of item response theory (IRT), such methods include comparison of item parameters (e.g., Lord, 1980; Thissen, Steinberg & Wainer, 1988, 1993) or of areas between item response functions (e.g., Raju, 1990). Efforts to understand why some examinees respond differentially to these items, however, are most often done outside the context of IRT (e.g., Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, & Cahalan, 2000). Mixture IRT models have recently been suggested as a useful means of investigating how qualitative examinee differences, such as differences in use of cognitive strategies (Mislevy & Verhelst, 1990) or use of different problem- solving strategies, may lead to differences in responses to test items (Embretson & Reise, 2000). Recent work with a mixture nominal model has demonstrated that it is possible to detect different latent groups in multiple choice data and to obtain diagnostic information about examinees based on their response patterns (Bolt, Cohen & Wollack, 2001).

Mixture models in IRT assume that while an IRT model holds for an entire population of respondents, different parameter values may apply for different latent classes of examinees. Examinees in a mixture model can be characterized by both a latent ability parameter as well as by a parameter indicating latent class membership. Interpreting differences in responses is aided by the discrete nature of the latent classes. Mixture IRT models also provide a mechanism for modeling DIF, albeit with respect to latent populations of examinees as opposed to manifest ones (e.g., as defined by characteristics such as ethnicity or gender). The mixture model used in this study employs constraints on item parameter estimates identified in a prior DIF analysis to define



BEST COPY AVAILABLE

latent groups of examinees distinguished by ethnic group-related nuisance dimensions.

## Methods

**Data.** Data for this study were taken from the 1996 NAEP State Mathematics Assessment for Grade 8. Items in this program were divided into 13 unique blocks, each requiring 15 minutes for completion. Each assessed student received a booklet containing three of the 13 blocks according to a design that ensured that each block was administered to a representative sample of students within each jurisdiction (Allen, Jenkins, Kulich & Zelenak, 1997). For this study, data were used from block 4. There were 21 multiple-choice items in this block. Each of these items measured achievement in one of five content areas, number and operation (7 items), measurement (4 items), geometry (4 items), data analysis including statistics and probability (2 items), and algebra and functions (4 items). In addition, a set of background variables were also available for each examinee. These included ethnicity, gender, mother's education, father's education, attitudes toward mathematics (e.g., "I like Math"), confidence in mathematics ability ("I am good at Math", "Understand most of math classes"), extent of belief about usefulness of mathematics (e.g., "Math is useful for solving everyday problems").

Two random sample of examinees were drawn from 18,958 students who took this block. There were 17,708 Caucasian, 2,859 African-American, and 1,250 Hispanic students. Sample 1 consisted of 1,000 Caucasian and 1,000 African-American examinees. Sample 2 consisted of a different random sample of 1,000 Caucasian examinees and 1,000 Hispanic examinees. That is, two sets of DIF comparisons were made in this study: One set between the Caucasian and African-American examinees in Sample 1 and a second set between the Caucasian and Hispanic examinees in Sample 2.

**DIF Detection.** A likelihood ratio test (Thissen, et al., 1988, 1993) was used to

BEST COPY AVAILABLE

detect DIF, between the two groups in each sample, one item at a time. This form of the LR test for DIF was conducted using the computer program MULTILOG (Thissen, 1991). Two models are estimated in this test, a compact model in which no DIF is assumed and an augmented model in which DIF is assumed to be possible in the studied item. In the compact model, the item parameters are constrained to be equal in the two groups. In the augmented item, the parameters for the studied item are unconstrained and the item parameters for the remaining 20 items are constrained to be equal in the two groups. The same compact model is used for all items. A different augmented model is estimated for each studied item. The LR test for the studied item consists of calculating the ratio of the likelihoods for the compact and augmented model. This likelihood ratio is asymptotically distributed as a $\chi^2$ with degrees of freedom equal to the difference in the number of parameters estimated in the augmented and compact models (Thissen, et al, 1993), and has been found to maintain Type I error control at the item level (Cohen, Kim, & Wollack, 1996). All DIF comparisons were assessed at the .05 level.

---

Insert Tables 1 and 2 about Here

---

Results of the DIF analyses are presented in Tables 1 and 2 for Sample 1, the Caucasian and African-American DIF comparisons, and for Sample 2, the Caucasian and Hispanic comparisons, respectively. In Sample 1, four DIF items were identified between Caucasian and African-American groups, Items 5 (Data Analysis), 11 (Number and Operations), 15 (Algebra and Functions), and 21 (Geometry). One of these items, Item 5, was harder for African-American examinees; the remainder were harder for Caucasian examinees. In Sample 2, three DIF items were identified between Caucasian and Hispanic examinees in Sample 2, Items 12 (Number and Operations), 14 (Algebra



BEST COPY AVAILABLE

and Functions), and 15 (Algebra and Functions). Items 12 and 14 were harder for Hispanic examinees and Item 15 was harder for Caucasian examinees. Neither the pattern of DIF results nor the content of the DIF items provided much information regarding the possible causes of the DIF in these two samples.

Next, the patterns of detected DIF and non-DIF items (shown in Tables 1 and 2) were used to set constraints in the computer program MULTILOG for a single two-group analysis of item parameters first in Sample 1 and then in Sample 2. The constraints in Sample 1 held non-DIF items equal and DIF items unconstrained between Caucasian and African-American groups. The resulting item parameter estimates from this two-group analysis are reported in Table 1. A similar set of two-group DIF analyses were done for Sample 2 as shown in Table 2. The resulting estimates are presented on the right side of Table 1 and 2, for Sample 1 and Sample 2, respectively. These were used to fix the values of the item parameter estimates in the two separate mixture IRT model analyses for Sample 1 and for Sample 2.

**The Mixture 3-Parameter Logistic Model (M3plM).** The IRT models for the items used in this study are given in Allen, et al., (1997). The items in block 4 were all calibrated using a 3-parameter logistic model (3plM). A separate mixture 3-parameter logistic model (M3plM) was used for the purpose of investigating latent classes of examinees associated with the ethnic group-related DIF that was detected in each of the two samples.

The M3plM is simple extension of a mixture Rasch model described by Rost (1990) in which a population of examinees is assumed to be composed of a fixed number of discrete latent classes. The IRT models are assumed to hold for each class but the item difficulty, discrimination, and guessing parameters differ for the different classes. Each examinee is parameterized both by a class membership parameter ($g = 1, ..., G$) and a within class ability parameter ($\theta_g$).

The probability of a correct response in a M3plM is modeled as

$$P(U = 1|g, \theta_g) = c_{ig} + (1 - c_{ig})\frac{\exp\{a_{ig}(\theta_g - \beta_{ig})\}}{1 + \exp\{a_{ig}(\theta_g - \beta_{ig})\}}, \quad (1)$$

where $\alpha_{ig}$ and $\beta_{ig}$ are the discrimination and location parameter, respectively, for item $i$ in class $g$, and $\theta_g$ is the trait parameter in class $g$.

Parameters called mixing proportions, $\pi_g$, are specified in the model to indicate the proportion of examinees in each class. These mixing proportions are constrained so that $\Sigma_{g=1}^{G}\pi_g = 1$. Class parameters $\mu_g$ and $\sigma_g$ denote the mean and standard deviation for $\theta_g$, respectively, for class $g$. These parameters permit associations between class and ability level.

**Estimation of Model Parameters.** A Markov-Chain Monte Carlo (MCMC) estimation algorithm employing Gibbs' sampling was used to estimate the model parameters and latent class memberships of examinees. This algorithm is implemented in the WINBUGS software (Spiegelhalter, Thomas, & Best, 2000). The MCMC algorithm samples a class membership for each examinee at each stage of the chain and then samples item and class parameter values conditional on those class memberships. An examinee's estimated class membership is determined by the proportion of times the examinee is sampled into each class over the course of the Markov chain. The algorithm can be used in this way to learn about both the classes as well as the class memberships of examinees.

The following priors were used for both M3plM analyses:

$$\theta_{gh} \sim Normal(\mu_g, 1), \quad h = 1, \ldots, N$$

$$g_h \sim \text{Bernoulli}(\pi_1, \pi_2), \quad h = 1, \ldots, N$$

$$\mu_g \sim Normal(0, 1), \quad g = 1, 2$$

$$(\pi_1, \pi_2) \sim Dirichlet(0.5, 0.5),$$

where $N$ is the total number of examinees. In addition, the item parameters were fixed at the values obtained from the final MULTILOG run (and shown in Table 1). These priors resulted in posterior distributions for which Gibbs' sampling could be done. In this study, $\sigma_g$ was fixed to 1 in both classes.

Starting values are needed for each parameter to define the first state of the Markov chain. All the starting values for the model parameters were randomly generated using the WINBUGS software. The WINBUGS code used for the M3plM is given in the appendix.

Determination of suitable burn-in was based on results from a chain run for a length of 11,000 iterations. Separate MCMC analyses were conducted for Sample 1 and for Sample 2. The chains for $\pi_1$, $\pi_2$, $\mu_1$, and $\mu_2$ for Sample 1 are illustrated in Figure 1. The chains for these same parameter estimates for Sample 2 are shown in Figure 2.

---

Insert Figures 1 and 2 About Here

---

WINBUGS provides several indices which can be used to help determine an appropriate length for the burn-in. These indices suggested that burn-in lengths of less than 100 iterations were reasonable. A conservative estimate of 1,000 iterations for the burn-in was used in Sample 1 and Sample 2, resulting in chains of 10,000 iterations used to obtain estimates of group membership and class size.

## Results

**Composition of Latent Groups.** Since the data for the analyses in Sample 1 consisted of 1,000 Caucasians and 1,000 African-Americans, the expectation was that, if the latent classes correspond directly to the two ethnic groups, the proportions in

the two latent classes would each be .5 and they would contain members of a single ethnic group. Results summarized in Table 3 indicate that this was not the case. The two-class solution using the M3plM yielded mixing proportions of approximately 67.5 percent in Class 1 and 32.5 percent in Class 2 (see Table 3a). That is, 67.5 percent of the individuals in the sample responded in a manner consistent with the Caucasian-like (C-like) pattern of DIF and 32.5 percent responded in a manner consistent with the African-American-like (AA-like) pattern of DIF. Substantial cross-over occurred in classification of Caucasian and African-American examinees into the latent class opposite to their ethnicity: 154 Caucasians were placed into Class 2, the AA-like group, and 503 African-American were placed into Class 1, the C-like group. The percent of examinees classified in the latent class consistent with their ethnic group was 67.2 It is evident from the data in Table 3 that the composition of the two latent classes clearly does not adhere to the ethnicity makeup of the sample.

---

Insert Table 3 About Here

---

A similar result was observed in Sample 2 (see Table 3b). The M3plM solution yielded mixing proportions of approximately 66.2 percent in Class 1, the C-like group, and 33.8 percent in Class 2, the Hispanic-like (H-like) group: 540 Hispanic examinees were classified in the C-like class, and 216 Caucasian were placed into the H-like class. In addition, 62.2 percent of the examinees were classified into classes that were consistent with their ethnic group membership.

**Comparison of ICCs for Manifest and Latent Groups.** An important assumption in this study is that members of a latent class are more alike in their responses and different from members of another latent group. These differences, furthermore, are more meaningful in explaining possible causes of DIF than are differences between

members of an ethnic group. To illustrate this, we compared ICCs for manifest groups (i.e., ethnic groups) with ICCs for latent groups. This set of comparisons helps to illustrate whether the pattern of differences between manifest groups is similar to the pattern between latent groups. Item parameter estimates for the manifest groups are given at the right side of Tables 1 and 2 for Samples 1 and 2, respectively.

To develop similar item parameter estimates for the latent groups, we used the classification results from the MCMC analyses to identify latent groups of C-like and AA-like. Item parameters next were estimated using the computer program MULTILOG and applying the same item-level constraints on DIF and non-DIF items as were used to obtain the item parameter estimates shown in Table 1. Finally, we equated the item parameter estimates for each of the two latent classes to the estimates in Table 1, using the test characteristic curve method via the computer program EQUATE (Baker, 1990). This placed the item parameter estimates for the two latent groups onto the respective metrics of the estimates for the two manifest groups for Sample 1. Similarly, item parameter estimates for C-like and H-like groups were obtained and placed on the same metric of the estimates for the Caucasian and Hispanic groups (in Table 2).

The ICCs for the 4 DIF items found between Caucasian and African-American examinees are shown in Figure 3 for both the manifest and latent groups. The ICCs for manifest groups are on the left, those for the latent groups are on the right. It is clear that the ICCs for the C-like and AA-like groups, however, show much greater differences than do those for the Caucasian and African-American examinees. As an example, the ICCs for latent groups on item 5 and 11 show C-like students have a very high probability of answering correctly.

---

Insert Figure 3 About Here

---

Figure 4 shows the similar ICC plots for the 3 DIF items for Caucasian vs Hispanic groups and C-like vs H-like groups. The left side graphs are for Caucasian and Hispanic groups and the right side ones are for C-like and H-like groups. The same patterns as seen in Figure 3 are present in Figure 4. This is particularly evident for Items 12 and 14, for which the differences between ICCs for C-like and H-like groups are substantially larger than for the manifest Caucasian and Hispanic groups.

Insert Figure 4 About Here

**Ability Differences Between Manifest and Latent Groups.** In each two-class solution, the abilities of the 2,000 examinees in the respective samples were estimated in the MCMC runs based on the final 10,000 iterations. Analyses of estimated ability between Caucasian (M = .96, SD = .99) and African-American (M = -.02, SD = 1.01) examinees indicated a clear difference in favor of Caucasian examinees. The difference between the C-like class (M = .60, SD = 1.09) and the AA-like class (M = .20, SD = 1.11), however, was much smaller. The difference between ability estimates for Caucasian (M = .73, SD = .97) and Hispanic (M = .07, SD = 1.00), however, was about half that between the C-like (M = .76, SD = .83) and H-like (M = -.30, SD = 1.05) classes. The reason for this was not immediately clear.

**Background and Attitude Differences Between Manifest and Latent Groups.** Results from the analyses of the eight background variables available on the sample of 18,958 examinees are presented in Tables 4 and 5 for Sample 1 and Sample 2, respectively. WINBUGS was again used with item parameters fixed at the values estimated from MULTILOG based on the results of the DIF analyses on Sample 1 and Sample 2 to identify latent groups for all examinees for each two-class solution. In addition, the mixing proportions and mean abilities for latent groups were fixed at the values

obtained from the initial analyses described above. In these WINBUGS runs, 6,000 iterations with a burn-in of 500 iterations were used. Results based on the final 5,500 iterations were used to obtain estimates of group membership.

---

Insert Tables 4 and 5 About Here

---

As shown in Table 4, of the 14,849 Caucasian students in Sample 1, 12,549 students were classed in the C-like group and 2,300 students in the AA-like group. About 50 percent of the 2,859 African-American students in Sample 1 were classified into the C-like class. Slightly more male African-American students were classified in the C-like class than African-American female students. Analysis of background information revealed more AA-like students indicated their parents' education levels were low (either "Didn't Finish High School" or "Graduate from High School"). More Caucasian and African-American students were classified into the C-like class, however, when the education levels were high ("Some Education after High School" or "Graduate from College").

Two-way ANOVAs were used to compare the characteristics between manifest and latent groups. Significant differences were found between Caucasian and African-American groups in math preference, math confidence, extent of belief about math usefulness, and extent of understanding in math class as shown in Table 7. African-American students had lager scores on each of these variable. The effect sizes, however, were relatively small. Significant differences were observed for the latent groups on all five variables. The C-like group had larger scores on each variable.

About 50 percent of African-American students were classed into the C-like group. African-American students indicated they felt they were better at mathematics ("I am Good at Math") than Caucasian students, but C-like students indicated this same



BEST COPY AVAILABLE

difference over AA-like students. Clearly, many of the African-American students who were classed into the C-like group were ones who felt better about their proficiency at mathematics.

---

Insert Tables 6 and 7 About Here

---

Of the 14,849 Caucasian students in Sample 2 (see Table 5), 11,606 students were classified into the C-like and 3,243 students into the H-like class. Of the total of 1,250 Hispanic students, 670 were classified into the C-like class and 580 into the H-like class, as shown in Table 5. There was a clear tendency for examinees in the C-like class to report their parents had higher education levels than examinees in the H-like class.

As can be seen in Table 6, significant differences were found between Caucasian and Hispanic examinees on "Discuss Studies at Home," and "I am Good at Math." Differences were found between latent groups, however, on all the five variables. The effect sizes for these results were moderately large.

## Conclusion

DIF occurs because of the intrusion of one or more nuisance dimensions on the measurement of the ability of interest (Ackerman, 1992; Roussos & Stout, 1996). In the present study, a mixture IRT model, the M3plM, was described and used to discretize this nuisance dimension in order to determine whether it would be possible to learn what might be responsible for the DIF that was detected. The use of manifest ethnic categories to identify those affected by ethnic group DIF does little to tell us about these causes. It appears at best to be a weak clue to the causes of ethnic group DIF. In other words, those examinees for whom items were actually functioning differentially were not accurately characterized by their ethnicity, despite the fact that the groups



BEST COPY AVAILABLE

were defined according to a prior ethnic group-based DIF analysis in two different samples.

Results of this study are consistent with previous work on gender DIF (Cohen & Bolt, 2002) and suggest that use of manifest group distinctions for detection of DIF do not clearly identify those examinees for whom the item truly functions differentially. Items which functioned differentially against Caucasian examinees did not do so for all Caucasians. Likewise, those items found to function differentially for or against African-Americans did not do so for all African-Americans. Similarly, items which functioned against Hispanic examinees did not do so for all Hispanic examinees. Thus, blanket identification of the groups affected as being distinguished by labels associated with manifest characteristics such as ethnicity may be a mistake in efforts to understand the root causes of DIF.

Patterns of ethnic group DIF were used in this study to define latent groups on the 1996 NAEP Grade 8 State Mathematics Assessment. In one sample, a DIF analysis was used to detect detected between Caucasian and African-American examinees. In a second sample, DIF was detected between Caucasian and Hispanic students. Latent groups were then defined for these same examinees based on conformation to the item parameters estimated based on the pattern of DIF in the ethnic groups. Comparison of memberships in the latent ethnic-like groups with membership in the manifest ethnic groups indicated that only about 67 and 62 percent of the examinees in Samples 1 and 2, respectively, were assigned to the same latent group as their ethnic group would indicate. An examination of the characteristics of examinees in the manifest and latent groups revealed ability differences between manifest groups for Sample 1 that were larger than between the latent groups. The reverse was found for Sample 2.

Latent ethnic-like groups showed more significant differences on background and attitude variables compared to examinees in the manifest ethnic groups. Even when

significant differences were observed between manifest groups for some variables, the effect sizes were smaller than for latent groups. These results point to clear differences between latent and manifest groups. Not surprisingly, members of each latent group were more alike than members of each manifest group.

Results of this study suggest that at least some of the explanation of DIF may be related to attitudes toward mathematics. This appears to explain some of the ethnicity-related DIF patterns better than ethnicity itself. Additional information about characteristics of latent groups would be helpful in explaining possible causes of ethnicity-related DIF. Results from this study may be specific to the NAEP samples analyzed. However, comparisons among the background and attitude variables clearly indicate the weak relationship between manifest group distinctions and causes of DIF. An advantage of the mixture model approach is that it can help identify potentially important examinee differences, thereby, providing further insight into how DIF affects the responses of different latent subpopulations of examinees.

BEST COPY AVAILABLE

**References**

Angoff, W. H. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–23). Hillsdale, NJ: Erlbaum.

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*(1), 67-91.

Allen, N.L., Jenkins, F., Kulick, E., & Zelenak, C.A. (1997). *Technical report of the NAEP 1996 State Assessment Program in Mathematics.* Washington, D.C.: National Center for Education Statistics.

Baker, F. B. (1990). *EQUATE* [computer program]. Madison, WI: University of Wisconsin, Laboratory of Experimental Design.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (April, 2001). *A mixture item response model for multiple-choice data.* Paper presented at the annual convention of the American Educational Research Association, Seattle, WA.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001b). *Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints.* Madison, WI: Center for Placement Testing, University of Wisconsin.

Borsboom, D. Mellenbergh, G. J., & Heerden, J. (2002). Different kinds of DIF: A Distinction between absolute and relative forms of measurement invariance and bias *Applied Psychological Measurement. 26,* 433–450.

Cohen, A. S., & Bolt, D. M. (2002). *A Mixture Model Analysis of Differential Item Functioning.* Paper presented at the annual convention of the American Educational Research Association, New Orleans, LA.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mawhah, NJ: Lawrence-Erlbaum.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology, 75*, 165-190.

Lord, F. M. (1980). *Applications of item response theory*. Hillsdale, NJ: Lawrence-Erlbaum.

Mislevy, R. J., & Verhelst, N. (1990). Modelling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195-215.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197–207.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

Roussos, L. & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement, 20*, 355-371.

Thissen, D. (1991). *MULTILOG* [computer program]. Chicago, IL: Scientific Software.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147–169). Hillsdale, NJ: Erlbaum.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Erlbaum

Samejima, F. (1972). A general model for free response data. *Psychometrika Monograph Supplement*, No. 18.

Samejima, F. (1997). Graded response model. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory*, pp. 85-100). New York:

BEST COPY AVAILABLE

Springer-Verlag.

Spiegelhalter, D., Thomas, A. & Best, N. (2000). *WINBUGS version 1.3* [computer program]. Robinson Way, Cambridge CB2 2SR, UK: MRC Biostatistics Unit, Institute of Public Health. Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG* [computer program]. Chicago, IL: Scientific Software.

Table 1. DIF Test Results for Sample 1: Caucasian and African-American Examinees

| | | | 3PL Item Parameter Estimates | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Estimates for Each Augmented model | | | | | | Final Estimates Based on DIF Pattern | | | | | |
| | | | Caucasian | | | African-American | | | Caucasian | | | African-American | | |
| Items | LR | Content[a] | a | b | c | a | b | c | a | b | c | a | b | c |
| 1 | 6.9 | N | 0.21 | -5.06 | 0.20 | 0.35 | -3.78 | 0.20 | 0.20 | -6.60 | 0.00 | 0.20 | -6.60 | 0.00 |
| 2 | 4.8 | M | 0.25 | -4.17 | 0.20 | 0.48 | -1.89 | 0.19 | 0.39 | -2.78 | 0.00 | 0.39 | -2.78 | 0.00 |
| 3 | 0.8 | G | 0.28 | -2.40 | 0.20 | 0.37 | -1.72 | 0.20 | 0.31 | -2.69 | 0.00 | 0.31 | -2.69 | 0.00 |
| 4 | 1.8 | N | 1.08 | -0.19 | 0.17 | 0.90 | -0.42 | 0.15 | 0.86 | -0.75 | 0.00 | 0.86 | -0.75 | 0.00 |
| 5 | 11.7* | D | 1.05 | -0.02 | 0.17 | 0.85 | 0.27 | 0.20 | 0.93 | -0.43 | 0.01 | 0.87 | 0.17 | 0.20 |
| 6 | 0.4 | N | 1.09 | -0.31 | 0.17 | 1.30 | -0.30 | 0.13 | 1.09 | -0.61 | 0.00 | 1.09 | -0.61 | 0.00 |
| 7 | 6.1 | G | 0.63 | 0.37 | 0.23 | 0.46 | 0.62 | 0.20 | 0.72 | 0.59 | 0.30 | 0.72 | 0.59 | 0.30 |
| 8 | 4.5 | M | 0.58 | 0.41 | 0.16 | 0.61 | 0.08 | 0.15 | 0.48 | -0.33 | 0.00 | 0.48 | -0.33 | 0.00 |
| 9 | 1.9 | N | 0.52 | 0.14 | 0.17 | 0.57 | -0.05 | 0.16 | 0.46 | -0.54 | 0.00 | 0.46 | -0.54 | 0.00 |
| 10 | 2.1 | A | 0.93 | 0.19 | 0.19 | 1.13 | 0.30 | 0.17 | 0.94 | -0.01 | 0.09 | 0.94 | -0.01 | 0.09 |
| 11 | 22.8* | N | 1.13 | 0.03 | 0.19 | 0.89 | -0.61 | 0.17 | 0.96 | -0.42 | 0.05 | 0.79 | -0.96 | 0.03 |
| 12 | 3.6 | N | 1.19 | 1.28 | 0.23 | 1.37 | 1.48 | 0.24 | 1.34 | 1.21 | 0.23 | 1.34 | 1.21 | 0.23 |
| 13 | 2.9 | N | 0.53 | 1.85 | 0.13 | 0.93 | 1.45 | 0.13 | 0.46 | 1.32 | 0.00 | 0.46 | 1.32 | 0.00 |
| 14 | 2.9 | A | 0.90 | 0.81 | 0.17 | 1.41 | 0.88 | 0.21 | 1.01 | 0.71 | 0.17 | 1.01 | 0.71 | 0.17 |
| 15 | 25.9* | A | 0.74 | 1.62 | 0.16 | 0.75 | 1.24 | 0.25 | 0.57 | 1.20 | 0.05 | 0.87 | 1.16 | 0.27 |
| 16 | 2.3 | D | 0.60 | 2.77 | 0.14 | 0.91 | 2.84 | 0.19 | 0.68 | 2.71 | 0.17 | 0.68 | 2.71 | 0.17 |
| 17 | 2.5 | G | 0.76 | 1.86 | 0.18 | 1.01 | 1.53 | 0.15 | 0.68 | 1.54 | 0.12 | 0.68 | 1.54 | 0.12 |
| 18 | 2.1 | M | 1.47 | 2.49 | 0.15 | 2.88 | 2.50 | 0.16 | 1.76 | 2.37 | 0.16 | 1.76 | 2.37 | 0.16 |
| 19 | 4.2 | M | 1.33 | 2.65 | 0.15 | 8.38 | 2.12 | 0.13 | 1.17 | 2.50 | 0.12 | 1.17 | 2.50 | 0.12 |
| 20 | 6.2 | A | 1.61 | 2.06 | 0.21 | 0.55 | 2.68 | 0.15 | 1.07 | 1.96 | 0.18 | 1.07 | 1.96 | 0.18 |
| 21 | 10.7* | G | 0.81 | 2.60 | 0.19 | 0.74 | 2.51 | 0.09 | 0.75 | 2.49 | 0.18 | 0.56 | 2.52 | 0.06 |

[a]N = Number & Operation, M = Measurement, G = Geometry, D = Data Analysis, A = Algebra & Function

*$p(\chi^2(df = 3) > 7.81) < .05$



BEST COPY AVAILABLE

Table 2. DIF Test Results for Sample 2: Caucasian and Hispanic Examinees

| | | | 3PL Item Parameter Estimates | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Estimates for Each Augmented Model | | | | | | Final Estimates Based on DIF Pattern | | | | | |
| | | | Caucasian | | | Hispanic | | | Caucasian | | | Hispanic | | |
| Items | LR | Content[a] | a | b | c | a | b | c | a | b | c | a | b | c |
| 1 | 4.6 | N | 0.21 | -5.40 | 0.20 | 0.31 | -4.37 | 0.20 | 0.21 | -6.63 | 0.00 | 0.21 | -6.63 | 0.00 |
| 2 | 3.8 | M | 0.25 | -4.40 | 0.20 | 0.42 | -2.36 | 0.19 | 0.37 | -3.26 | 0.00 | 0.37 | -3.26 | 0.00 |
| 3 | 5.9 | G | 0.28 | -2.78 | 0.20 | 0.34 | -1.74 | 0.19 | 0.33 | -2.59 | 0.00 | 0.33 | -2.59 | 0.00 |
| 4 | 0.5 | N | 1.06 | -0.56 | 0.18 | 0.99 | -0.61 | 0.19 | 0.97 | -0.76 | 0.14 | 0.97 | -0.76 | 0.14 |
| 5 | 1.0 | D | 1.05 | -0.36 | 0.17 | 0.99 | -0.29 | 0.21 | 1.01 | -0.43 | 0.18 | 1.01 | -0.43 | 0.18 |
| 6 | 0.0 | N | 1.08 | -0.66 | 0.18 | 1.17 | -0.63 | 0.13 | 1.04 | -0.92 | 0.00 | 1.04 | -0.92 | 0.00 |
| 7 | 6.8 | G | 0.61 | 0.01 | 0.22 | 0.64 | -0.39 | 0.17 | 0.52 | -0.65 | 0.07 | 0.52 | -0.65 | 0.07 |
| 8 | 3.4 | M | 0.58 | 0.07 | 0.16 | 0.80 | 0.11 | 0.21 | 0.53 | -0.46 | 0.00 | 0.53 | -0.46 | 0.00 |
| 9 | 4.3 | N | 0.52 | -0.19 | 0.17 | 0.61 | -0.43 | 0.14 | 0.49 | -0.80 | 0.00 | 0.49 | -0.80 | 0.00 |
| 10 | 0.0 | A | 0.91 | -0.17 | 0.18 | 0.88 | -0.26 | 0.14 | 0.77 | -0.60 | 0.00 | 0.77 | -0.60 | 0.00 |
| 11 | 3.7 | N | 1.11 | -0.36 | 0.18 | 0.96 | -0.59 | 0.19 | 0.96 | -0.64 | 0.16 | 0.96 | -0.64 | 0.16 |
| 12 | 10.4* | N | 1.20 | 0.93 | 0.23 | 1.59 | 1.11 | 0.21 | 1.18 | 0.86 | 0.23 | 1.56 | 1.05 | 0.21 |
| 13 | 1.2 | N | 0.55 | 1.49 | 0.13 | 0.71 | 1.47 | 0.12 | 0.46 | 1.13 | 0.00 | 0.46 | 1.13 | 0.00 |
| 14 | 9.6* | A | 0.90 | 0.46 | 0.17 | 1.67 | 0.89 | 0.29 | 0.77 | 0.14 | 0.07 | 1.78 | 0.85 | 0.30 |
| 15 | 11.5* | A | 0.74 | 1.26 | 0.16 | 0.60 | 1.17 | 0.21 | 0.60 | 1.00 | 0.08 | 0.64 | 1.17 | 0.23 |
| 16 | 2.2 | D | 0.59 | 2.43 | 0.14 | 0.50 | 2.96 | 0.18 | 0.53 | 2.59 | 0.16 | 0.53 | 2.59 | 0.16 |
| 17 | 3.6 | G | 0.75 | 1.51 | 0.18 | 0.84 | 1.13 | 0.14 | 0.62 | 1.18 | 0.10 | 0.62 | 1.18 | 0.10 |
| 18 | 2.0 | M | 1.44 | 2.15 | 0.15 | 1.59 | 2.34 | 0.15 | 1.37 | 2.17 | 0.14 | 1.37 | 2.17 | 0.14 |
| 19 | 0.7 | M | 1.41 | 2.29 | 0.15 | 1.31 | 2.14 | 0.14 | 1.13 | 2.25 | 0.13 | 1.13 | 2.25 | 0.13 |
| 20 | 3.4 | A | 1.59 | 1.72 | 0.21 | 1.40 | 1.77 | 0.17 | 1.36 | 1.65 | 0.18 | 1.36 | 1.65 | 0.18 |
| 21 | 2.6 | G | 0.80 | 2.27 | 0.19 | 0.57 | 2.05 | 0.10 | 0.39 | 2.00 | 0.02 | 0.39 | 2.00 | 0.02 |

[a]N = Number & Operation, M = Measurement, G = Geometry, D = Data Analysis, A = Algebra & Function

*$p(\chi^2(df = 3) > 7.81) < .05$



BEST COPY AVAILABLE

Table 3. Manifest versus Latent Class Assignment

Table 3a. Sample 1: Caucasian & African-American

| Latent Class | Ethnic Group | | |
|---|---|---|---|
| | Caucasian | African-American | Total |
| C-like | 846 | 503 | 1349 |
| AA-like | 154 | 497 | 651 |
| Total | 1000 | 1000 | 2,000 |

Table 3b. Sample 2: Caucasian & Hispanic

| Latent Class | Ethnic Group | | |
|---|---|---|---|
| | Caucasian | Hispanic | Total |
| C-like | 784 | 540 | 1324 |
| H-like | 216 | 460 | 676 |
| Total | 1000 | 1000 | 2,000 |

BEST COPY AVAILABLE

Table 4. Characteristics of Manifest & Latent Groups in Sample 1: Caucasian & African-American Examinees

| | | Ethnic Groups | | | |
|---|---|---|---|---|---|
| | | Caucasian | | African-American | |
| | | C-like | AA-like | C-like | AA-like |
| Gender | Male | 6188 | 1090 | 657 | 639 |
| | Female | 6361 | 1210 | 773 | 790 |
| Mother's Education Level | Didn't Finish High School | 900 | 360 | 102 | 149 |
| | Graduate from High School | 3285 | 772 | 404 | 454 |
| | Some Edu. after High Sch. | 2617 | 288 | 332 | 202 |
| | Graduate from College | 4566 | 480 | 399 | 376 |
| | I don't Know | 1137 | 382 | 186 | 242 |
| | Omitted | 44 | 18 | 7 | 6 |
| Father's Education Level | Didn't Finish High School | 869 | 335 | 95 | 121 |
| | Graduate from High School | 2945 | 672 | 361 | 380 |
| | Some Edu. after High Sch. | 2117 | 274 | 242 | 141 |
| | Graduate from College | 4801 | 461 | 317 | 283 |
| | I don't Know | 1804 | 551 | 414 | 502 |
| | Omitted | 13 | 7 | 1 | 2 |
| Discuss Studies at Home | Almost Everyday 4[a] | 5320 | 853 | 551 | 550 |
| | Once or Twice a Week 3[a] | 3800 | 575 | 444 | 381 |
| | Once or Twice a Month 2[a] | 1243 | 237 | 139 | 110 |
| | Never or Hardly ever 1[a] | 2103 | 581 | 274 | 316 |
| | omitted | 83 | 54 | 22 | 72 |
| Agree? I Like Math | Strongly Agree 5[a] | 1968 | 234 | 390 | 304 |
| | Agree 4[a] | 5081 | 735 | 568 | 535 |
| | Undecided 3[a] | 2711 | 522 | 231 | 235 |
| | Disagree 2[a] | 1686 | 406 | 134 | 165 |
| | Strongly Disagree 1[a] | 1042 | 378 | 100 | 158 |
| | omitted | 61 | 25 | 7 | 32 |
| Agree? I am Good at Math | Strongly Agree 5[a] | 2330 | 208 | 416 | 265 |
| | Agree 4[a] | 6266 | 777 | 646 | 569 |
| | Undecided 3[a] | 2473 | 594 | 214 | 287 |
| | Disagree 2[a] | 1014 | 436 | 105 | 177 |
| | Strongly Disagree 1[a] | 356 | 250 | 39 | 94 |
| | omitted | 110 | 35 | 10 | 37 |
| Agree? Math is Useful for Solving Everyday Problems | Strongly Agree 5[a] | 4827 | 849 | 708 | 610 |
| | Agree 4[a] | 5284 | 830 | 489 | 437 |
| | Undecided 3[a] | 1448 | 295 | 121 | 160 |
| | Disagree 2[a] | 598 | 144 | 58 | 94 |
| | Strongly Disagree 1[a] | 227 | 103 | 21 | 37 |
| | omitted | 165 | 79 | 33 | 91 |
| Agree? Understand Most of Math Classes | Strongly Agree 5[a] | 3157 | 299 | 448 | 324 |
| | Agree 4[a] | 6824 | 1083 | 719 | 697 |
| | Undecided 3[a] | 1447 | 455 | 135 | 203 |
| | Disagree 2[a] | 721 | 250 | 78 | 111 |
| | Strongly Disagree 1[a] | 259 | 147 | 27 | 42 |
| | omitted | 141 | 66 | 23 | 52 |
| Total | | 12549 | 2300 | 1430 | 1429 |

[a]Assigned Scores to Compare Means between Groups

BEST COPY AVAILABLE

Table 5. Characteristics of Manifest & Latent Groups in Sample 2: Caucasian & Hispanic Examinees

| | | Ethnic Groups | | | |
|---|---|---|---|---|---|
| | | Caucasian | | Hispanic | |
| | | C-like | H-like | C-like | H-like |
| Gender | Male | 5724 | 1554 | 326 | 290 |
| | Female | 5882 | 1689 | 344 | 290 |
| Mother's Education Level | Didn't Finish High School | 782 | 478 | 144 | 165 |
| | Graduate from High School | 2950 | 1107 | 155 | 133 |
| | Some Edu. after High Sch. | 2466 | 439 | 104 | 60 |
| | Graduate from College | 4360 | 686 | 121 | 61 |
| | I don't Know | 1009 | 510 | 143 | 158 |
| | Omitted | 39 | 23 | 3 | 3 |
| Father's Education Level | Didn't Finish High School | 747 | 457 | 111 | 126 |
| | Graduate from High School | 2667 | 950 | 128 | 119 |
| | Some Edu. after High Sch. | 1986 | 405 | 97 | 48 |
| | Graduate from College | 4577 | 685 | 129 | 79 |
| | I don't Know | 1617 | 738 | 203 | 206 |
| | Omitted | 12 | 8 | 2 | 2 |
| Discuss Studies at Home | Almost Everyday 4[a] | 4969 | 1204 | 241 | 190 |
| | Once or Twice a Week 3[a] | 3546 | 829 | 204 | 138 |
| | Once or Twice a Month 2[a] | 1150 | 330 | 73 | 66 |
| | Never or Hardly ever 1[a] | 1873 | 811 | 146 | 147 |
| | omitted | 68 | 69 | 6 | 39 |
| Agree? I Like Math | Strongly Agree 5[a] | 1869 | 353 | 124 | 66 |
| | Agree 4[a] | 4759 | 1057 | 265 | 199 |
| | Undecided 3[a] | 2495 | 738 | 149 | 133 |
| | Disagree 2[a] | 1528 | 564 | 80 | 74 |
| | Strongly Disagree 1[a] | 898 | 522 | 49 | 90 |
| | omitted | 57 | 29 | 3 | 18 |
| Agree? I am Good at Math | Strongly Agree 5[a] | 2249 | 289 | 96 | 45 |
| | Agree 4[a] | 5852 | 1191 | 329 | 184 |
| | Undecided 3[a] | 2232 | 835 | 146 | 158 |
| | Disagree 2[a] | 874 | 576 | 68 | 110 |
| | Strongly Disagree 1[a] | 301 | 305 | 25 | 56 |
| | omitted | 98 | 47 | 6 | 27 |
| Agree? Math is Useful for Solving Everyday Problems | Strongly Agree 5[a] | 4456 | 1220 | 248 | 186 |
| | Agree 4[a] | 4899 | 1215 | 291 | 227 |
| | Undecided 3[a] | 1343 | 400 | 76 | 71 |
| | Disagree 2[a] | 549 | 193 | 31 | 34 |
| | Strongly Disagree 1[a] | 208 | 122 | 14 | 17 |
| | omitted | 151 | 93 | 10 | 45 |
| Agree? Understand Most of Math Classes | Strongly Agree 5[a] | 3013 | 443 | 147 | 79 |
| | Agree 4[a] | 6320 | 1587 | 354 | 269 |
| | Undecided 3[a] | 1306 | 596 | 95 | 110 |
| | Disagree 2[a] | 621 | 350 | 50 | 55 |
| | Strongly Disagree 1[a] | 218 | 188 | 15 | 36 |
| | omitted | 128 | 79 | 9 | 31 |
| Total | | 11606 | 3243 | 670 | 580 |

[a] Assigned Scores to Compare Means between Groups

BEST COPY AVAILABLE

Table 6. Attitude-Related Characteristics of Manifest & Latent Groups in Sample 1

| | Caucasian | | African-American | | Effect | C-like | | AA-like | | Effect |
|---|---|---|---|---|---|---|---|---|---|---|
| Agree? | Mean (SD) | N | Mean (SD) | N | Size* | Mean (SD) | N | Mean (SD) | N | Size* |
| Discuss Studies at Home | 2.95 (1.12) | 14712 | 2.88 (1.15) | 2765 | | 2.98 (1.10) | 13874 | 2.80 (1.20) | 3603 | 0.16 |
| I Like Math | 3.36 (1.18) | 14763 | 3.59 (1.22) | 2820 | -0.19 | 3.45 (1.16) | 13911 | 3.19 (1.28) | 3672 | 0.22 |
| I Am Good At Math | 3.64 (1.01) | 14704 | 3.72 (1.08) | 2812 | -0.08 | 3.76 (0.96) | 13859 | 3.27 (1.17) | 3657 | 0.49 |
| Math is Useful | 4.10 (0.95) | 14605 | 4.20 (0.98) | 2735 | -0.10 | 4.14 (0.92) | 13781 | 4.03 (1.08) | 3559 | 0.12 |
| Understand Most of Math Class | 3.89 (0.93) | 14642 | 3.95 (0.94) | 2784 | -0.06 | 3.97 (0.89) | 13815 | 3.63 (1.05) | 3611 | 0.37 |

* $p < 0.01$



BEST COPY AVAILABLE

Table 7. Attitude-Related Characteristics of Manifest & Latent Groups in Sample 2

| | Caucasian | | Hispanic | | Effect | C-like | | H-like | | Effect |
|---|---|---|---|---|---|---|---|---|---|---|
| Agree? | Mean (SD) | N | Mean (SD) | N | Size* | Mean (SD) | N | Mean (SD) | N | Size* |
| Discuss Studies at Home | 2.95 (1.12) | 14712 | 2.76 (1.18) | 1205 | 0.17 | 3.00 (1.09) | 12202 | 2.75 (1.21) | 3715 | 0.22 |
| I Like Math | 3.36 (1.18) | 14763 | 3.34 (1.21) | 1229 | | 3.45 (1.14) | 12216 | 3.05 (1.25) | 3776 | 0.34 |
| I Am Good At Math | 3.64 (1.01) | 14704 | 3.37 (1.08) | 1217 | 0.27 | 3.76 (0.94) | 12172 | 3.17 (1.12) | 3749 | 0.60 |
| Math is Useful | 4.10 (0.95) | 14605 | 4.05 (0.97) | 1195 | | 4.12 (0.92) | 12115 | 4.02 (1.05) | 3685 | 0.11 |
| Understand Most of Math Class | 3.89 (0.93) | 14642 | 3.72 (1.00) | 1210 | | 3.98 (0.88) | 12139 | 3.55 (1.05) | 3713 | 0.47 |

* $p < 0.01$



BEST COPY AVAILABLE

## Appendix: WINBUGS Code Used for Mixture 3PL Model

```
model
{
 # 3PL Mixture model analysis of Ethnic group DIF items on 1996
 # NAEP Math of grade 8, block 4. DIF items between Caucasian and
 # African-American were found with Likelihood Ratio tests using
 # Multilog. This model uses item parameter estimates obtained on a
 # second run of Multilog with constraints based on the DIF and
 # non-DIF items.

  for (j in 1:N) {
   for (k in 1:T) {
         r[j,k]<-resp[j,k]
}}
  for (j in 1:G) {
  alpht[j]<-alph[j]
  }

#  3PL model
   for (j in 1:N) {
       for (k in 1:T) {
           tt[j,k]<- exp(-a[gmem[j],k]*(theta[j] - b[gmem[j],k]))
           p[j,k]<- c[gmem[j],k]+(1-c[gmem[j],k])/(1 + tt[j,k])
           r[j,k]~dbern(p[j,k])
           }
         theta[j] ~ dnorm(mut[gmem[j]],1)
         gmem[j] ~ dcat(pi[1:G])
        }

#  Priors
   pi[1:2]~ ddirch(alpht[1:2])
   mut[1] ~ dnorm(0.,1.)
   mut[2] ~ dnorm(0.,1.) I(,mut[1])

# Fixed item parameters from Multilog

  a[1,1] <-  0.20
  a[2,1] <-  0.20
  a[1,2] <-  0.39
  a[2,2] <-  0.39
       :
  a[1,20] <- 1.07
  a[2,20] <- 1.07
  a[1,21] <- 0.75
```

```
  a[2,21] <- 0.56

  b[1,1]<-  -6.60
  b[2,1]<-  -6.60
  b[1,2]<-  -2.78
  b[2,2]<-  -2.78
        :
  b[1,20]<-  1.96
  b[2,20]<-  1.96
  b[1,21]<-  2.49
  b[2,21]<-  2.52

  c[1,1] <-  0.00
  c[2,1] <-  0.00
  c[1,2] <-  0.00
  c[2,2] <-  0.00
        :
  c[1,20] <- 0.18
  c[2,20] <- 0.18
  c[1,21] <- 0.18
  c[2,21] <- 0.06
}

# Data
list(N=2000, T=21, G=2, alph=c(5,5),
resp=structure(.Data=c(
1,1,1,1,1,1,0,1,1,1,1,0,1,0,0,0,0,0,1,0,0,
1,1,1,0,1,1,1,0,0,1,1,1,0,1,1,0,0,0,0,1,0,
1,1,0,1,1,1,1,1,0,1,0,1,0,0,0,0,1,0,0,0,0,
1,1,1,1,0,1,0,0,0,0,0,0,0,1,0,1,0,1,0,0,0,

:

1,1,1,1,1,1,0,0,0,1,1,0,0,0,0,1,1,0,1,0,0,
1,1,1,0,0,1,0,1,0,0,1,0,0,0,1,0,0,1,0,0,0,
1,0,1,0,0,0,1,0,0,0,0,0,1,0,0,0,0,1,0,1,0,
1,1,1,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0),
.Dim=c(2000,21)))
```

Figure 1: Sampling Histories for the First Two-Class Solution

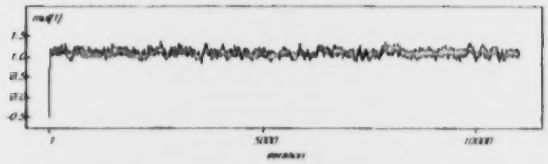


Figure 1a: Sampling History for $\mu(\theta_1)$.

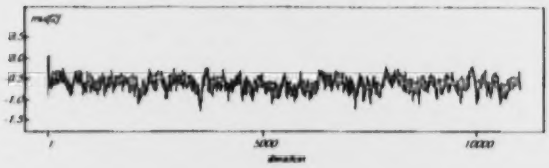


Figure 1b: Sampling History for $\mu(\theta_2)$.

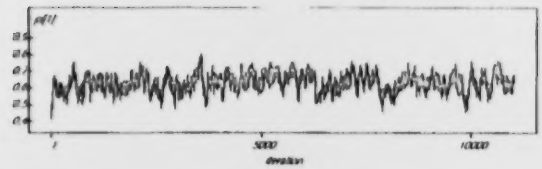


Figure 1c: Sampling History for $\pi_1$.

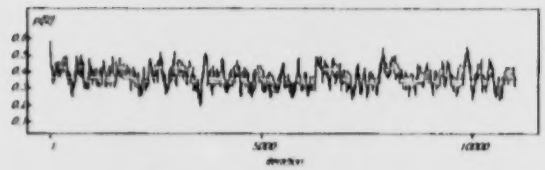


Figure 1d: Sampling History for $\pi_2$.

Figure 2: Sampling Histories for the Second Two-Class Solution

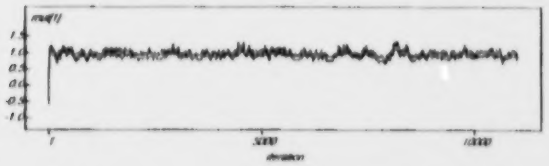


Figure 2a: Sampling History for $\mu(\theta_1)$.

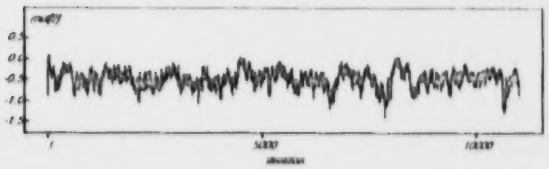


Figure 2b: Sampling History for $\mu(\theta_2)$.

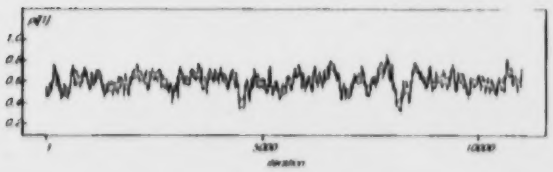


Figure 2c: Sampling History for $\pi_1$.

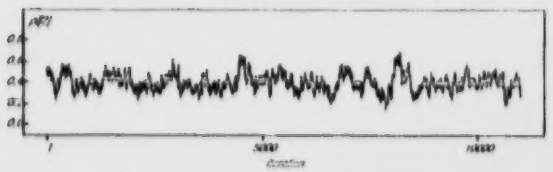
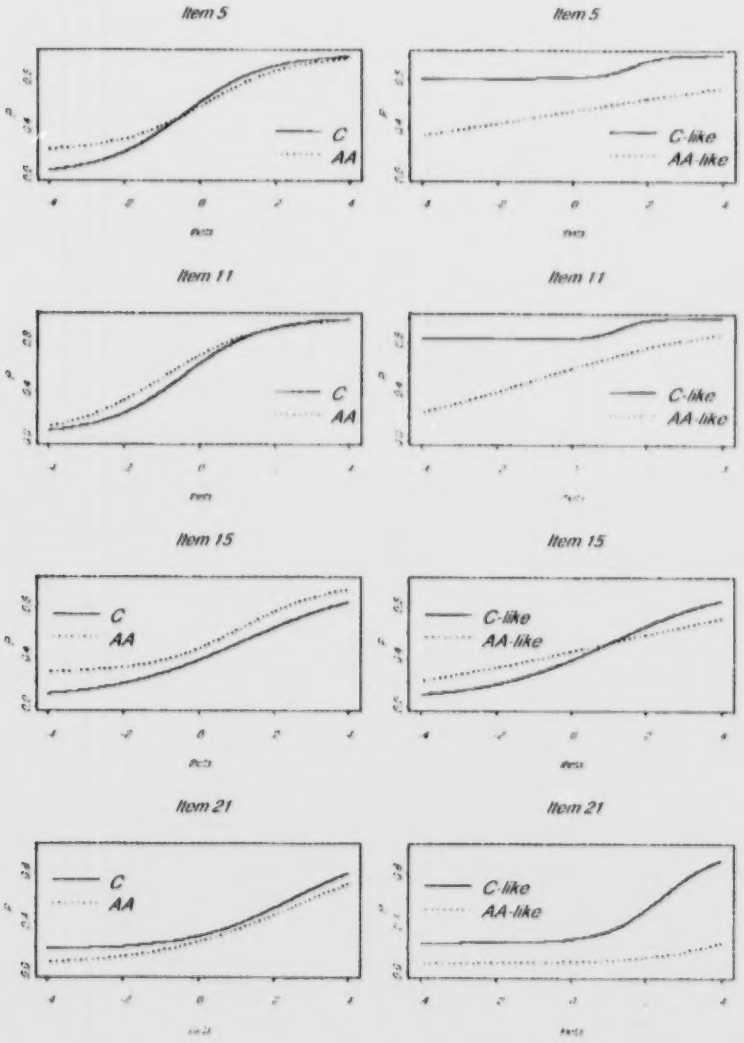


Figure 2d: Sampling History for $\pi_2$.



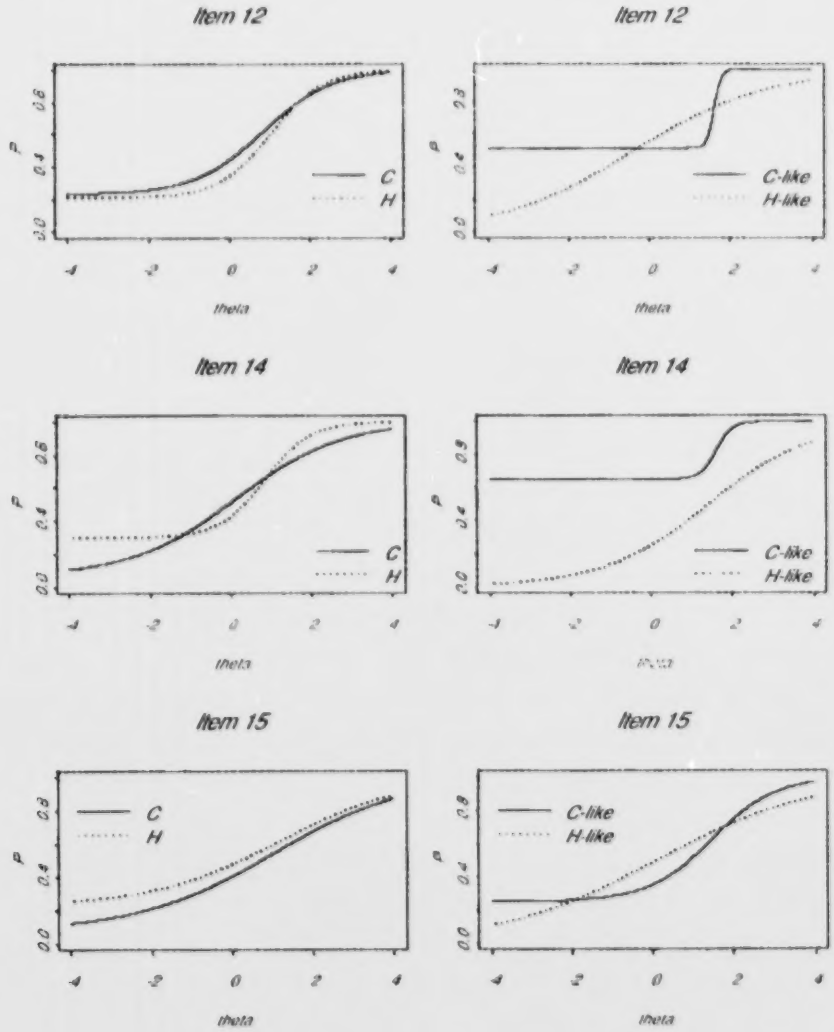
BEST COPY AVAILABLE

Figure 3: Item Characteristic Curves of DIF items For the First Two-Class Solution




BEST COPY AVAILABLE

Figure 4: Item Characteristic Curves of DIF items For the Second Two-Class Solution






BEST COPY AVAILABLE

TM034967

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

ERIC Educational Resources Information Center

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

| | |
|---|---|
| Title: A Mixture IRT Model Analysis of Ethnic Group DIF | |
| Author(s): Taehoon Kang and Allan S. Cohen | |
| Corporate Source: 2003 NCME Conference (Uni. of Wisconsin - Madison) | Publication Date: |

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ☐

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ☑

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

*I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.*

**Sign here, → please**

| | |
|---|---|
| Signature: | Printed Name/Position/Title: Taehoon Kang |
| Organization/Address: 1025 W. Johnson, #393 Madison WI 53706 | Telephone: 608-[illegible]-2848 FAX: 608-263-4291 |
| | E-Mail Address: taehoonkang@wisc.edu Date: 8/19/03 |

(Over)

## III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

| |
|---|
| Publisher/Distributor: |
| Address: |
| Price: |

## IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

| |
|---|
| Name: |
| Address: |

## V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: **University of Maryland**
**ERIC Clearinghouse on Assessment and Evaluation**
**1129 Shriver Lab, Bldg 075**
**College Park, MD 20742**
**Attn: Acquisitions**

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**University of Maryland**
**ERIC Clearinghouse on Assessment and Evaluation**
**1129 Shriver Lab, Bldg 075**
**College Park, MD 20742**
**Attn: Acquisitions**